ECEC 412/621: Computer Architecture
Solution Key for Quiz 3

Prof. Naga Kandasamy
ECE Department
Drexel University

June 1, 2016

1. Show the design of a four-way set associative write through cache of size 512 KB having 64-byte blocks. Also, show how the 30-bit main memory address is decomposed to access the cache. Assume that memory is byte addressable. What is the final size (including tags and status bits) in KB of the cache structure?

The 30-bit address is partitioned as follows:

- *Offset bits*: The least significant 6 bits (5–0) are used as the offset bits to locate the desired bytes within the cache block.
- *Index bits*: We have
$$2^{\text{index}} = \frac{512 \times 1024}{4 \times 64} = 2048.$$
So, 11 bits (16–6) are used as index bits to locate the set in which the memory block is stored.
- *Tag bits*: The most significant 13 bits (29–17) are used as the tag bits.

The final size of the cache includes, in addition to the cache blocks, the overhead due to the tag and status bits, which is $(64 \times 8 + 1 + 13) \times 2048 \times 4$ bits or 526 KB.

2. Answer the following questions related to caches.

(a) Set-associative caches typically have lower miss rates than direct-mapped ones. Why is it so? However, caches with large (8 or 16) set associativities are typically not built. Provide a reason why.

Caches with large associativities such as 8 or 16 are not built due to the following reasons:

1. The cache design is quite complex and large in terms of logic gates/circuits. This slows down the operation of the cache; that is, increased hit times.
2. Analysis on benchmark programs indicate that the miss rate does not improve that significantly when we go from a 4-way set associative cache to one with a higher associativity number.

(b) Can the random block replacement strategy be used in a direct mapped cache? Explain your answer clearly.

No. Since a memory address maps to a single location in the cache, we do not get the option of selecting a victim block to replace (like we do in set-associative caches).

3. Consider a two-level instruction cache hierarchy comprising direct-mapped level one and level two caches, L1 and L2, respectively. Suppose that in 1000 memory references, there are 30 misses in the L1 cache, and 10 misses in the L2 cache. Also, assume the following:

- Hit time in the L1 cache is 1 processor cycle.
- Hit time in the L2 cache is 4 processor cycles.
- Miss penalty incurred in the L2 cache is 50 processor cycles.

(a) What is the local miss rate for the L2 cache?

Miss $\text{rate}_{\text{L2}} = 10/30 = 0.33$.

(b) What is the average instruction access time for the above cache hierarchy?

Hit $\text{time}_{\text{L1}} = 1$ cycle.

Miss $\text{rate}_{\text{L1}} = 30/1000 = 0.03$.

Hit $\text{time}_{\text{L2}} = 4$ cycles.

Miss $\text{rate}_{\text{L2}} = 10/30 = 0.33$.

$$\begin{aligned}\text{AMAT} &= \text{Hit time}_{\text{L1}} + \text{Miss rate}_{\text{L1}} \times (\text{Hit time}_{\text{L2}} + \text{Miss rate}_{\text{L2}} \times \text{Miss penalty}_{\text{L2}}) \\ &= 1 + 0.03 \times (4 + 0.33 \times 50) \\ &\approx 1.6\end{aligned}$$